

2重音節マルコフモデルによる日本語の 文節音節認識候補の曖昧さの解消効果†

荒木 哲郎^{††} 村上 仁一^{††} 池原 悟^{††}

日本語の音声認識において、音声の物理的特性に着目した認識アルゴリズムで解決できない認識の曖昧さを、日本語の表現の持つ特徴に着目した方法で解消する問題について論じ、特に日本語の音節の連鎖特性を用いた音声認識装置の音節候補絞り込みの効果を明らかにする。音節の連鎖特性としては、日本語の単語または文節の音節数等の性質からみて、また0重、1重に比べて、2重マルコフ連鎖のエントロピーがかなり小さくなることから、本論文では2重マルコフ連鎖特性を取り上げその効果を実験により明らかにする。実験では、新聞記事70日分の音節連鎖特性を求め、これを用いて音節マトリックスで表された1,000文節の文節音節候補を対象（ただしマトリックスは、置換誤りのみを含むように模擬的に作成したもの）に、音節候補の絞り込み効果を求めた。その結果、マトリックスの各音節候補内に正しい音節が含まれる場合は、文節音節列10位までに、正解音節列が94.1~97.2（第1位は、60.4~66.3）%含まれ、さらに音節認識装置からの各音節候補の認識重み確率も用いた場合は、10位までの文節正解率は、95.8~98.7（第1位は72.4~78.8）%に向上することがわかった。また前者の正解率は1重マルコフ連鎖を用いた場合に比べて10位までで18.3~21.7（第1位で28.4~32.5）%高い値であり、2重マルコフ連鎖による絞り込み効果が著しく大きくなることを示している。

1. はじめに

日本語の連続音声認識の方法には、連続した音声をいくつかの音節の組合せ（単語等の単位）でパターンマッチングを取る方法^{1),2)}と、一つ一つの音節に分割（セグメンテーション）して組み立てる方法³⁾⁻⁵⁾がある。前者は音節単位の認識の曖昧さを、長単位化することで克服することを狙ったものであるが、この方法を任意の文章の入力に適用しようとする、膨大な数の（無限に近い）音節の組合せパターンの認識が必要となり現実的には実現が難しい。後者は認識の単位が比較的少数の音節数（100種程度）に分割されるため任意文の入力には適しているが、セグメンテーションを正しく行うのが難しいこと、および高い音節認識精度を得るのが難しいことと理由で、従来実現が困難であった。

本論文では、任意の日本語文章の入力を目指す上から、後者の方法を取り、言語処理を用いて音節の曖昧さを解消することを狙いととしている。すなわち、文節単位の区切って発声された音声の認識において、セグメンテーションは正しく行われる（音節区間の境界が

正しく認識される）と仮定したときに得られる文節の各音節認識候補を、相互に組み合わせて得られる音節列の中から、言語処理を用いて正しい文節の音節列を求める問題を研究する。

言語処理の方法は、日本語の文法的、意味的特性を応用するものであるが、一般的には膨大な量の候補が生じるため、処理量も膨大なものとなる。一方、日本語の文法的、意味的特性は、日本語の音節間の結合関係（音節連鎖特性）にも現れると考えられることから、本論文では日本語処理を応用した音節認識率向上策を検討する第1ステップとして、日本語の音節連鎖特性を用いた音節認識候補絞り込みの効果を明らかにする。

従来の音節、文字などの連鎖特性を用いた研究としては、英語の、文字認識の分野において認識誤りの文字を訂正するのに、2文字（1重マルコフ）、3文字（2重マルコフ）の連鎖特性を用いた研究⁶⁾、ならびに離散単語音声認識の分野で単語単位の2重マルコフ連鎖特性を用いた研究⁷⁾がある。

一方日本語においては、かな文字の認識や単音節音声認識に対して、かなの2文字連鎖確率特性を用いた研究⁸⁾、や2文字、3文字の連鎖パターンの有無を用いた（連鎖パターンが存在するか否かの離散的な情報を利用）研究^{10),11)}はあるが、3文字（2重マルコフ）以上の連鎖確率特性による効果を、定量的に調べたような事例は未だ見当たらない。これは2重マルコフ以上になると、解析アルゴリズムが複雑になり処理量が

† Effect of Reducing Ambiguity of Recognition Candidates in Japanese "Bunsetsu" Units by 2nd-order Markov Model of Syllables by TETSUO ARAKI, JINICHI MURAKAMI and SATORU KEHARA (Natural Language Processing Laboratory, NTT Communication and Information Processing Laboratories).

†† NTT 情報通信処理研究所自然言語処理研究部

増大することおよび、実際の日本文について信頼できる音節連鎖確率データを収集することが難しいことにあると思われる。そこで本論文では、前者に対しては文献7)にも示されているように、動的計画法特にマルコフモデル上の Viterbi アルゴリズムを、2重以上のマルコフ連鎖確率に適用すること(ただし、文献7)では1重マルコフに適用)、また後者については先の日本文音声出力の研究¹⁸⁾において実現された高精度な言語解析技術(単語分割処理、単語読み(音節)の付与処理技術など)を用いて、新聞記事(漢字かな混じり文)70日分を文節単位に区切り、さらに助詞(「は」「を」「へ」)の音節変換や連続音声に伴う長音化、鼻音化の処理などを行うことにより、大量の信頼できる音節連鎖確率が入手可能になることで、解決を図った。

本論文では、最初に新聞記事70日分を用いて、0重、1重、2重マルコフの音節連鎖のエントロピーを調べ、0重、1重マルコフの連鎖特性に比べて、2重マルコフ連鎖のエントロピーが小さくなることを示し、2重マルコフの音節連鎖確率が音節認識候補の絞り込みに非常に有効であるという推測を立てる。次に2重マルコフの音節連鎖を用いた音節候補の絞り込み効果を調べるために、日本文音声入力モデルを設定し、音響処理結果の表現法の一つである音節マトリックス/ラティス⁴⁾のシミュレーションによる実験方法について述べる。最後に実験結果と考察を述べる。

2. 日本文の文節構造と音節連鎖確率の特徴

2.1 文節構造タイプと音節論的構成

日本文における基本的な言語の単位として、本論文では、文節を次のように定義する。すなわち、一つ以上の自立語(詞)={名詞、動詞、形容詞、形容動詞、副詞、連体詞、接統詞、感動詞、形式名詞}と付属語(辞)={助詞、助動詞、接辞}から構成される一つの単位を文節と呼ぶ。一般に前者は必ず文節に含まれるが、後者は必ずしも含まれるとは限らない。

このような文節構造に基づき、自立語の品詞に着目した文節タイプの出現頻度を、新聞記事の8,198文節数に対して調べると、図1のような傾向となり、①名詞、②名詞+辞、③動詞、および④動詞+辞のタイプが多い(9割以上も占め

る)ことがわかる。

このような文節に対しては、音節数またはモーラ数の面から次のような特徴、「動詞はほとんどが和語であるのに対し、名詞は半数以上が漢語で、他に和語、外来語などを含むこと、さらにモーラ数の観点からは、名詞、動詞は3~4モーラが多いこと、また辞(助詞、助動詞)は1~2モーラが多い」がある¹⁵⁾。このような性質から考えると、大部分の文節のモーラ数を3~6モーラ付近と推察することができる(実際にわれわれの実験でも、4.1節で後述するように文節当りの平均音節数は6である)。

次に文節内に現れる音節列の配列法を決定する問題について考える。一般に日本語の文節として意味のある文節に構成されるためには、文節として現れる音節の列がランダムに構成されるのではなく、ある一つの拘束性(例えば撥音「ン」は文節の先頭にくることはあり得ないとか、「ア」の次に「ル」がくる確率が非常に高いこと等)に従って規定されると考えられる。そこでこのような拘束性として、ここでは文節内の各位置に現れる音節が、該音節の直前/前方位置の音節によってのみ規定される(一般にマルコフ性と呼ばれる)場合について考える。このとき問題は、文節内の音節列の決定に当たって、文節内の拘束性及ぶ範囲(マルコフ性の次数)を最低(一般に多いほど確定度が高い)どこまで考慮すれば有効であるかを求めることである。日本語には一般に自立語に種々の付属語がついて多様な表現をなすという特質があるため、一文節内で自立語と付属語にまたがって拘束性を考えるこ

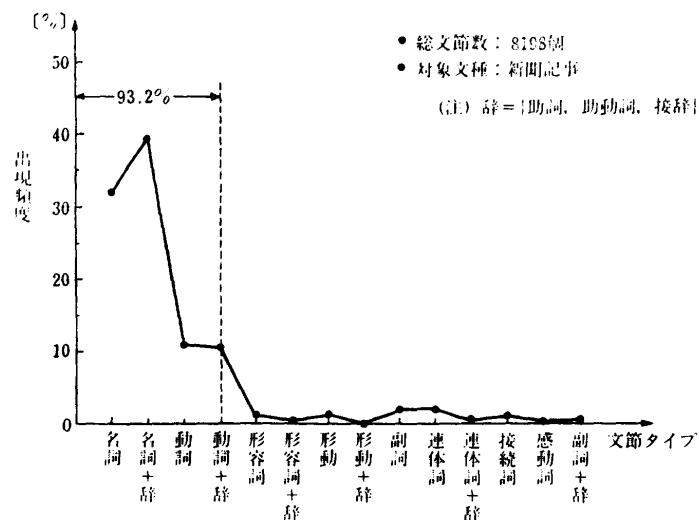


図1 文節構造タイプの出現頻度

Fig. 1 Occurrence ratio of "bunsetsu" type.

とはあまり得策ではない*こと、また付属語は種類が少なく文節境界の情報を利用することにより、比較的容易に決定され得ると考えられる。そこで文節内では自立語内の拘束性を基本と考えればよく、文節当りの平均音節数が6であること、および前記の名詞、動詞の音節数の特徴等から1重マルコフでは不十分で最低でも2重マルコフ以上が必要と推測され、文節における音節列の曖昧さを解消する効果を知る上で、2重マルコフが一つの大きな役割を担うと考えられる。

2.2 文節単位で求めたマルコフ音節連鎖確率によるエントロピー

前節で述べた文節内の拘束性は、音節の連鎖確率でとらえることができ、情報論的にはエントロピーの形で表される。

ここでは、新聞記事70日分を用いて、実際に文節単位中の音節の0重、1重、2重マルコフの音節連鎖確率を調べた。その際、先に開発した実験用の日本文音声出力システム¹⁰⁾を用いてべた書きの漢字かな混じり文を、文節切りした後で、助詞「は」、「を」、「へ」の音節変換や、連鎖音声に伴う長音化、鼻音化の処理などを行った。

ここで用いた音節の種類は、清音(ア, サ, 他44)濁音(ガ, ザ, 他18), 半濁音(バ, ピ, 他5), 拗音(キャ, ジャ, 他33), 鼻音(ガ行鼻音のみ, 8), 長音, 撥音(ン), 促音(ッ)から成る111音節である。また音節連鎖確率を求めるために使用した全音節数は、約290万個であり、文節当りの平均音節数は6.11であった。

これらの0重、1重、2重マルコフの音節連鎖確率^{**}を用いてエントロピー(各々、 H_0, H_1, H_2 と表す)を求めると次のようになる。

$$H_0 = -\sum P(i) \log_2 P(i) = 6.15$$

$$H_1 = -\sum P(i, j) \log_2 P(j|i) = 4.46$$

$$H_2 = -\sum P(i, j, k) \log_2 P(k|i, j) = 3.07$$

また文節単位に切り出し、前後を空白と見なして求

* 一般に自立語にどのような付属語がつかは、文内の他の文節間との関係等で決まることが多いため、本論文のように一文節内の範囲で考えるときは、自立語と付属語との拘束性は除外しても差し支えないと考えられる。

** m 重マルコフ連鎖確率は、現在の音節(i_m)の出現確率がそれ以前の m 個の音節列($i_{n-m}, i_{n-m+1}, \dots, i_{n-1}$)出現確率にだけ依存して定まり、条件付き確率 $P(i_m | i_{n-m}, i_{n-m+1}, \dots, i_{n-1})$ で表される。また $P(i, j, k, \dots)$ は同時確率を表す。本論文では、文節の境界は空白(U)とし、文節のあらゆる位置で各連鎖確率を定義可能とするため、例えば、文節先頭文字 x に対する2重マルコフ連鎖確率は、 $P(x|UU)$ と表すこととする。また空白文字(U)の扱い等によりエントロピーの値が、文献19)、20)の場合より大きくなっている。

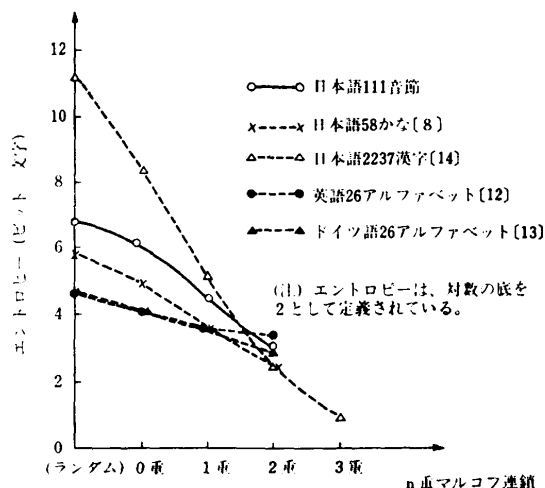


図2 日本語音節および他言語のエントロピー
Fig. 2 Entropies of Japanese syllable and the other printed languages.

めたここでのエントロピーは、従来文献9)で示したようなかな文字の任意な文(各文節がつながった場合)のエントロピー(0重:5.55, 1重:4.24)に比べて、大きくなっている。これは音節の種類(111個)がかな文字のセット(83個)より多いためである。

さらに漢字かな混じり文¹⁴⁾や外国文(英文¹²⁾, 独語文¹³⁾の場合のエントロピーと比較すると(文節/文単位等の条件は異なる), 図2のようになる。同図より、字種の少ない英文などに比べて日本語のカタカナ/音節などの場合は、エントロピーが大きくなる傾向にあること、また2重マルコフ付近で小さくなる傾向にあることがわかる。

また、1重、2重マルコフの音節連鎖確率を用いる時に要する記憶量の目安となるゼロ確率(先の脚注で述べた条件付確率がゼロである1重、2重の音節連鎖組)の割合は、1重では全体(112²)の約43%、2重では全体(112³)の約95%を占めており、高次になるにつれて使われる音節連鎖の組合せが非常に局所化されていることがわかる。

本論文では、上記に示したように文節単位の2重マルコフ音節連鎖確率のエントロピーが、従来知られている1重マルコフ音節連鎖のエントロピー等と比べてかなり小さくなる事実に着目し、その2重マルコフの音節連鎖確率が音節認識候補の絞り込みに非常に有効であるという推測をたて、以後その効果を実験的に把握していくことにする。

3. 日本文音声入力モデルと実験方法

2章で述べた0~2重のマルコフの音節連鎖確率を用いて、文節内の音節列の曖昧さをどの程度解消できるかを、定量的に把握するために、図3のような日本文音声入力モデルを考える。ここで同モデルにおける音声認識装置は、音声のセグメンテーションが誤りなく行われるものと仮定し、拗音、長音、鼻音、濁音、半濁音を含む111音節に対する認識候補を出力するものとする。

3.1 音節マトリックス

音声認識装置における音響処理の結果は一般に、文献4)のように音節認識の曖昧さ(置換誤り)と、音節区間の切り出し(セグメンテーション)の曖昧さ(脱落、挿入誤り)を表し得る音節系列の形式(音節ラティス)を用いて表すことができる。本論文では簡単のために、セグメンテーションの誤りが無い場合を考え、図4のように置換誤りのみから成る音節認識候補の形式(特にこれを音節マトリックスと呼ぶ)を対象とする。

またここでは簡単のため、音声認識装置から出力される音節認識候補の特性に基づき、次のような性質をもつコンフュージョン表をもとに(文献16)を参照し、

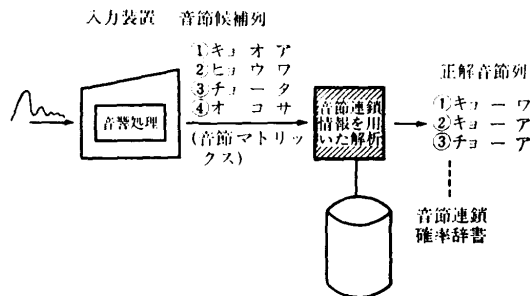


図3 日本文音声入力モデル
Fig. 3 Model of Japanese text input processing system by speech.

文節内位置(ma×20)

	1	2	3	4	5	6	7
第1候補	オ		ク	ラ	シ	ー	ワ
第2候補	ホ		フ	ガ	チ		ハ
第3候補	ホ		フ	ア			ハ
第4 "			ケ	ダ			ダ
第5 "				ハ			
第6 "				ハ			

図4 音節マトリックスの例(:「オークラショーワ」
…大蔵省は)

Fig. 4 An example of syllable matrix:
「オークラショーワ」.

文節単位の音節マトリックスを疑似的に作成する。

- ① 誤りは、正解音節と同じ母音をもつ音節間でのみ生じ、他の母韻の音節間では生じない(例えば、「ア」は「タ」や「パ」には誤るが、「イ」や「ク」などには誤らない)。
- ② 促音(「ッ」)、撥音(「ン」)および長音については、誤りを生じない(正解音節のみ)。またガ行鼻音化の誤り傾向は、ガ行の誤りの場合と同じ値を用いる。
- ③ 平均の音節候補数は、3.5であり、第1位候補(正解音節)に対する平均的な音節正解率*は、約89.0%である。

なお、今回の実験で用いる、文節単位の音節マトリックスシミュレーションでは、もとの正解音節が必ず音節マトリックスの中に含まれる場合を扱う。

3.2 最尤な音節認識候補の選出方法

音声認識装置から出力された複数個の音節認識結果の候補群(音節マトリックス)から、2.2節で述べた音節連鎖確率を用いて、最尤な音節認識候補列を見つける方法について述べる。特に、2重の音節マルコフ連鎖確率の場合は、展開された全候補系列に対して網羅的に調べると、一般には膨大な処理時間を要する。これを解決するために、動的計画法特にマルコフモデル上のViterbiアルゴリズム¹⁷⁾を、2重マルコフに拡張して行う方法について述べる。

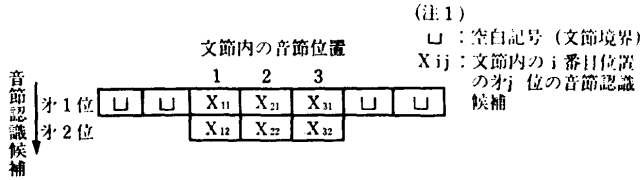
音節マトリックスから最尤な候補列を見つける問題にViterbiアルゴリズムを適用するには、文献17)のようにマルコフ源からの出力系列Xを、音声認識装置に入る音節列に、また雑音のある通信路を経て観測されるZを、音声認識装置から出力される曖昧な音節列に対応させればよい。また観測され得るすべてのZが、音節マトリックスから導かれ、17)で示されるように、事後確率P(X|Z)を最大にするような音節列が、trellisと呼ばれる一種の有向グラフ上の最短ルートとして求められる。

すなわち、 $X=(x_0, x_1, \dots, x_{n+1})$, $Z=(z_0, z_1, \dots, z_{n+1})$ に対して、

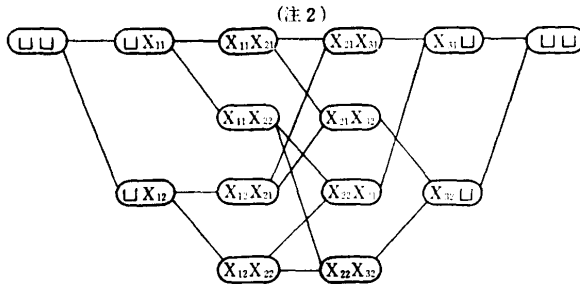
$$\Pr(x_0, \dots, x_{n+1} | z_0, \dots, z_{n+1}) = \frac{\Pr(z_0, \dots, z_{n+1} | x_0, \dots, x_{n+1}) \cdot \Pr(x_0, \dots, x_{n+1})}{\Pr(z_0, \dots, z_{n+1})} \quad (1)$$

となり、またm重マルコフ源、記憶なしの通信路の

* 各音節認識候補には音声認識処理の結果を表す認識の正しさを示す認識重み確率がそれぞれ付与されているものとする。



音節マトリックスに、2重(一般にm重)マルコフ連鎖確率を適用して、最尤なパス(音節列)を見つける問題は、m個前の可能な各音節列を、一つの状態とするようなトレリスを作成することにより、一重マルコフのViterbiアルゴリズムに帰着できる。



(注2) 各遷移枝には、それぞれ2重音節マルコフ連鎖確率(例えばノードX₁₁X₂₁とX₂₁X₃₁の間ではPr(X₃₁|X₁₁X₂₁))の負の対数値が割り当てられる。

図5 ビテルビアルゴリズムを2重マルコフ連鎖に拡張したときのトレリスの例

Fig. 5 An example of trellis used in Viterbi algorithm extended to 2nd-order Markov source.

時、(1)式の分子*は、

$$\prod_{i=1}^{n+m} \Pr(z_i | x_i) \cdot \Pr(x_i | x_{i-m}, \dots, x_{i-1}) \quad (2)$$

となる。分母は定数値をとるから、(1)式を最大とするには、(2)式を最大にすればよいことになる。また計算を容易にするため、(2)式の確率値の代わりに、その対数値を用い、(2)式と等価な(3)式を最小とする問題に置きかえることができる。

$$-\sum_{i=1}^{n+m} \ln \Pr(z_i | x_i) - \sum_{i=1}^{n+m} \ln \Pr(x_i | x_{i-m}, \dots, x_{i-1}) \quad (3)$$

ここで、Viterbi アルゴリズムは、一つ前の時刻の状態(音節文字 x_{i-1}) から次の時刻の状態 (x_i) に遷移する1重マルコフの有向グラフ (trellis と呼ぶ) の上で最小のパスを見つける方法を与えるものである。これをm重マルコフの場合に拡張するには、m個前の音節列の組合せ (x_{i-m}, \dots, x_{i-1}) を一つの状態とみな

* x_0 と x_{n+1} は空白を、 x_1, \dots, x_n は音節を表す。また(2)式で x_j, z_j が $j < 0$ の時は x_0, z_0 を、 $j > n$ の時は x_{n+1}, z_{n+1} を表すものとする。

して、直前の状態 (x_{i-m}, \dots, x_{i-1}) から直後の状態 (x_{i-m+1}, \dots, x_i) への遷移枝の割り当てを、①直前の状態を表す音節列で先頭を除く後半の $m-1$ 個の音節文字列が、直後の状態を表す音節列の先頭の $m-1$ 個の列に一致する時だけ、状態間に遷移枝を設定し、②その遷移枝に、(3)式の $\ln \Pr(x_i | x_{i-m}, \dots, x_{i-1})$ を割り当てることにより、m重マルコフの最小パスを見つける問題を、一重マルコフの Viterbi アルゴリズムに帰着できる。

図5に、音節マトリックスに2重マルコフ音節連鎖確率を用いた時の trellis の例を示す。同図で各音節位置のそれぞれのノードにおいて、そのノードに至るまでのルートの中で最小値(またはk位まで)のルート(音節部分列に該当)のみを、ずっと最後の音節位置まで保持して求めていくことにより、最短ルート(またはk位まで)の最尤音節列を求めることができる。

4. 実験結果と考察

4.1 実験対象の日本語と実験環境条件

音声入力の対象および音節連鎖確率データの収集に用いた日本語は、一般の新聞記事文(日経新聞の広告、株式等の文形式以外を除く)であり、次の条件を満たすものである。

- (i) 外国語や会話用のカギ括弧を含む文は除く。
- (ii) 文節単位に区切られ、長音、鼻音などを含む111音節を用いて表記される。
- (iii) 入力文節の総数は、1,000個。
- (iv) 入力1文節当りの平均音節数は、5.8~6.2
- (v) 音節連鎖確率データの収集には、約70日分のすべて新聞記事を用いた。またこの時の確率の収集に直接用いた同日付の新聞記事の中から入力対象の日本語を選ぶ時(標本内データと呼ぶ)と、それ以外の日付の新聞記事から入力を選ぶ時(標本データと呼ぶ)に分けて実験を行った。

4.2 0重、1重、2重の音節マルコフ連鎖確率による文節正解率

音声認識装置の曖昧な音節認識候補(音節マトリックス)の中から、0重、1重、2重の音節マルコフ連鎖確率を用いて最尤な文節単位の音節候補列を求めた時および、音節連鎖確率の情報に、さらに各音節に対する認識の正しさ(重み確率)を付加して求めた時

* 入力文には、約2%の分かち書きの誤り(文節切りの誤り一切れすぎなど)が含まれているが、今回の実験範囲ではそれによる影響は全体的な傾向を把握する上では少ないと判断した。

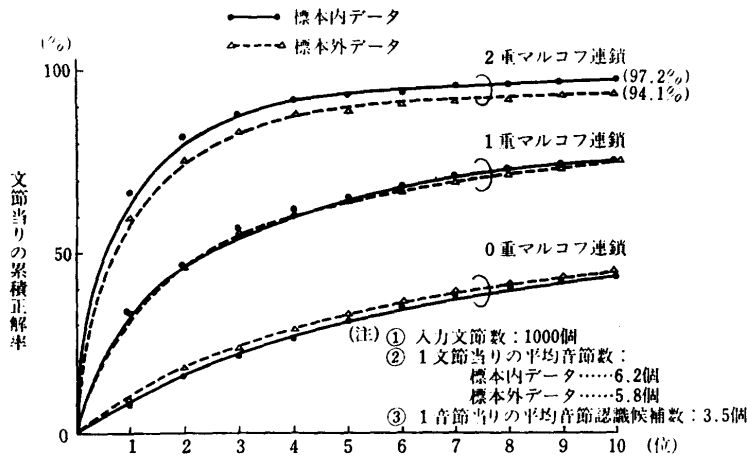


図6 音節連鎖確率の種類と文節当りの累積正解率 (音節認識重み確率未使用)
Fig. 6 Percentage of correct Japanese syllable strings in 1st-10th ranks based on 0th, 1st and 2nd Markov chain probabilities of syllables.

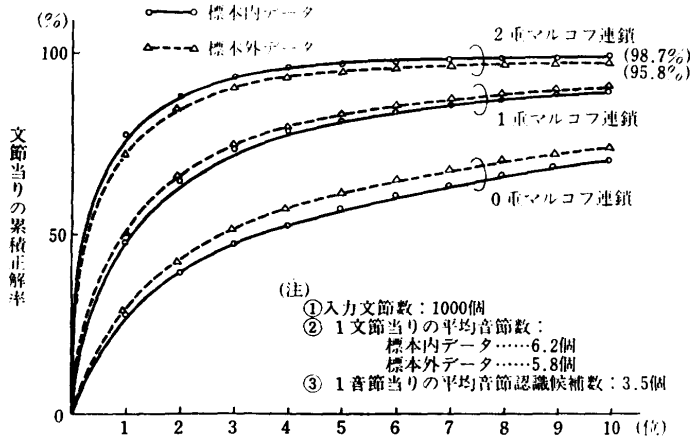


図7 音節連鎖確率の種類と文節当りの累積正解率 (音節認識重み確率使用)
Fig. 7 Percentage of correct Japanese syllable strings in 1st-10th rank based on 0th, 1st and 2nd Markov chain probabilities, of syllables in addition to probability of correct syllable using acoustic processing.

の、10位内文節正解率の実験結果*をそれぞれ図6および図7に示す。同図より次のことがわかる。

- ① 2重音節マルコフ連鎖のみを使用する場合は、文節音節列第10位までに、正解音節列が94.1% (標本外データのとき)~97.2% (標本内データのとき)**含まれており、特に第1位内でも60.4~

* 2重音節マルコフ連鎖確率を用いた時に、不正解となった文節の例、ならびに不正解の文節数が入力文節の音節数にはほぼ比例して高くなっている**。

** 一般には、標本内データの場合の方が標本外データの場合よりも正解率は高いが、標本量の増加に伴い両者は接近する。実際の真の値はその中間にあると考えられるため、正解率を以後標本外データの場合(小さい方)と、標本内データの場合(大きい方)の両者の正解率で示すことにする。

66.3%含まれている。これは0重マルコフの場合に比べて、10位内で47.8~54.1% (第1位で51.9~57.0%) 向上し、1重マルコフに比べて10位内で18.3~21.7% (第1位で28.4~32.5%) 向上している。

- ② さらに音声認識装置から出力される各認識候補の重み確率情報を併用すると、2重マルコフ連鎖の場合で、10位までの文節正解率は95.8~98.7% (第1位は72.4%~78.8%)**に向上する。これは0重マルコフに比べて10位内で21.8~28.2% (第1位で43.3%~50.9%)、1重マルコフに比べて10位内で5.1~9.3% (第1位で22.1~29.6%) だけ高い値となっている。

また、2重音節マルコフ連鎖確率の場合について、各音節の認識重み確率0.89の値を、0.69, 0.81, 0.92と変化させた時の、文節正解率の変化(第1位および第10位)を、図8に示す。ここで同図において2重マルコフ音節連鎖確率により、音声認識装置からの曖昧さを、どの程度解消できたかを以下に示す。一般に、各音節当りの正解率(認識重み確率に相当)を r 、文節の音節数を L とすると、音声認識装置(音響処理の結果)から出力される文節当りの音節列正解率は、 $p=r^L$ で表される。今文節内の平均音節数を、 $L=5.8$ とする時、 r に対する文節正解率曲線は図8の点線のように求まる。この点線の文節正解率

曲線を、2重マルコフの音節連鎖確率情報を用いることにより、どの程度文節正解率を向上できるか(曖昧さの解消効果)を、図8のハッチングによる矢印で表すことができる(文節正解率を、20~60%程度向上)。これは2重マルコフ連鎖確率情報を用いることにより、音節認識候補の曖昧さを大幅に解消できることを示している。

次に0重, 1重, 2重のマルコフ連鎖のエントロピーと文節正解率の関係を調べるための簡単な試みとして、ここでは音節表記の日本語における各音節の出現分布関数を仮定して、実験で求めた第1位の文節正解

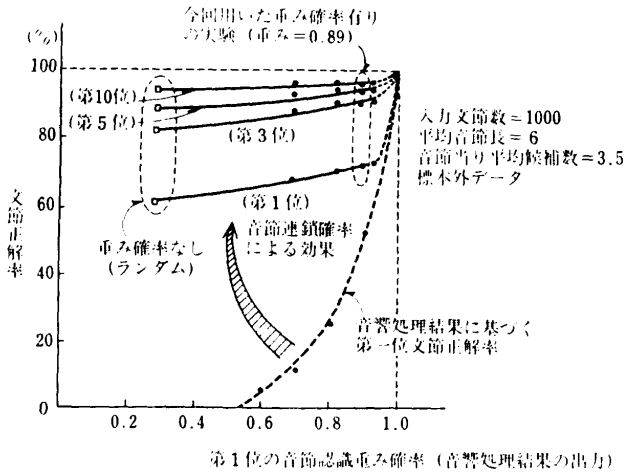


図 8 2重マルコフ音節連鎖確率による文節音節認識候補の曖昧さの解消効果
 Fig. 8 Effect of reducing ambiguity on bunsetsu syllable recognition candidates by 2nd-order Markov probabilities of syllables.

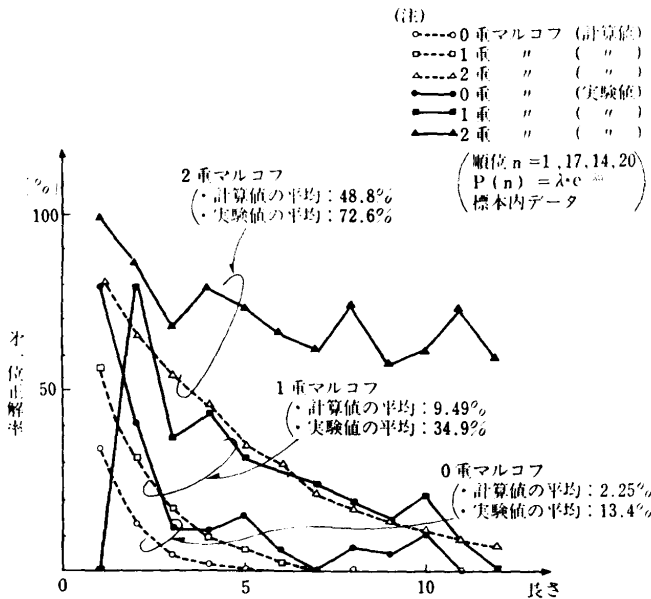


図 9 第1位文節正解率の実験値と計算値の関係
 Fig. 9 Relation between experimental results and theoretical values for 1st "bunsetsu" correct rate.

率を、前記のエントロピーから推定する方法を、付録1に示す。同付録1の計算式(A.6)を用いることにより、各エントロピー $H_i (i: 0, 1, 2)$ に対する第1位の文節正解率(計算値)が求まり、これと前に実験で求めた第1位の文節正解率(実験値)との関係を図9に示す。同図より、近似式(A.6)を用いてエントロピーから推定される文節正解率は、かなり変動があり妥当

な表現方法とは言い難いが、エントロピーと文節正解率の関係をj知る上で一つの目安を与える。

4.3 0重, 1重, 2重マルコフ音節連鎖確率の組合せ効果と足切り効果

ここでは2重マルコフ連鎖確率を用いた時の効果を、さらに向上させる手段として次のものを考える。

- (1) 2重マルコフ連鎖確率において確率が0(候補が棄却される)の音節を、他音節の2重マルコフ連鎖確率か、または同じ音節の1重または0重マルコフ連鎖確率で置き換えることにより、棄却候補の音節中から本来の正しいものを救済する方法。
- (2) 2重マルコフ連鎖確率において出現頻度の小さい確率の音節(棄却される可能性が多い)を足切りして(確率を0とする)、その部分を(1)のように2重, 1重または0重マルコフ連鎖確率で置き換えて救済する方法。

(1)については前述したように、2重マルコフ連鎖確率ではゼロ確率のものが全体の95%を占め、使用される確率の分布が非常に局所化されているために、本来日本語として存在するが、たまたま今回収集した標本データの中にこのような音節の組合せが含まれていない可能性がある。またこの場合については、1重マルコフ連鎖確率が存在する時と存在しない時、すなわち2重, 1重マルコフ連鎖確率 p_2, p_1 に対して、

- (i) $p_2(x_3|x_1, x_2)=0$ かつ $p_1(x_3|x_2) \neq 0$
 - (ii) $p_2(x_3|x_1, x_2)=0$ かつ $p_1(x_3|x_2)=0$
- の2通りの場合がある。これらに対して、

- ① 標本内データの場合のように常に元の正しい文節候補の音節列に対する2重マルコフ連鎖確率が存在する時には、2重マルコフ連鎖確率のみを用いる方が最善であり、1重マルコフ連鎖確率を用いるとかえって逆効果になる。
- ② 標本外データのように必ずしも元の正しい音節列に対して、2重マルコフ連鎖確率が存在するとは限らない時、次善の策として上記(1)のように他音節の2重マルコフまたは同音節の1重, 0重マルコフ連鎖確率を用いて上記のような救済効果

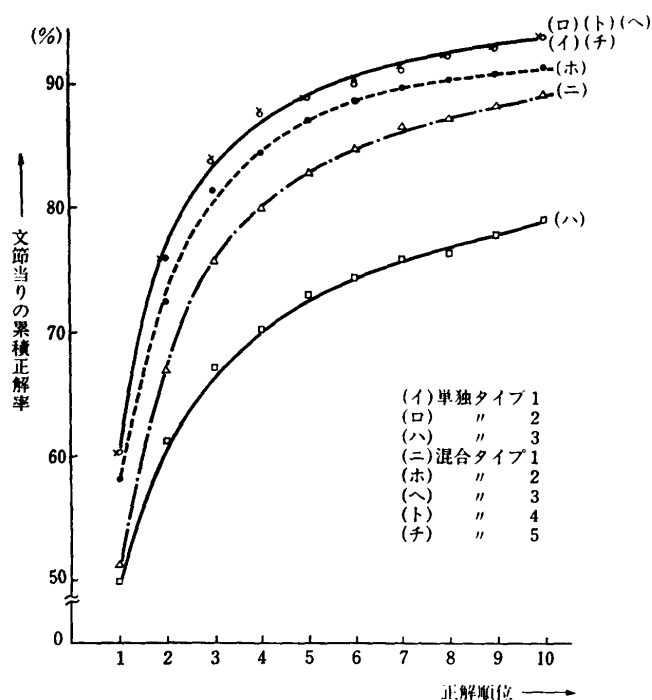


図 10 0重, 1重, 2重マルコフ連鎖確率の組合せによる合成効果
 Fig. 10 Composite effect of using combinations of the 0th, 1st and 2nd Markov chain probabilities of syllables.

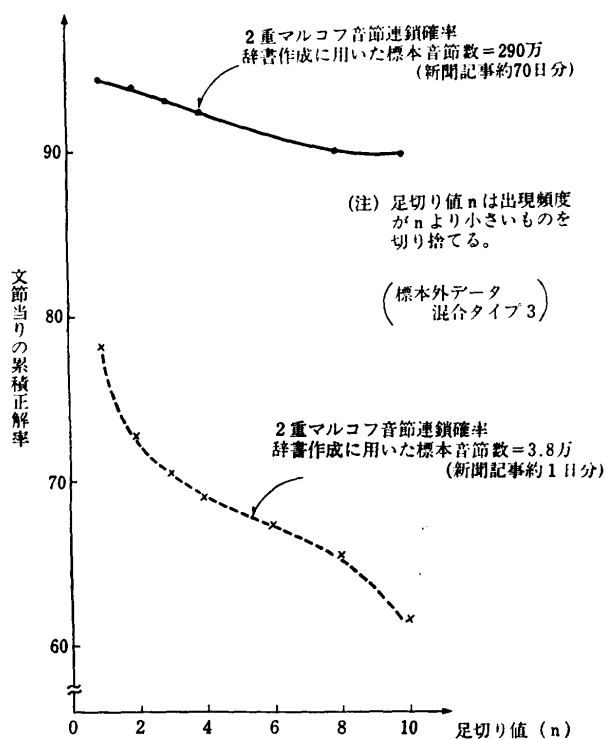


図 11 低頻度出現値の2重マルコフ音節連鎖確率を足切りした時の効果
 Fig. 11 Effect of neglecting the 2nd-order Markov chain probabilities which have low occurrence frequencies.

を期待できる見込みがある。

特に②の場合について、付録2に示すような0重、1重、2重マルコフ連鎖確率の代表的ないくつかの組合せタイプに対して実験を行い、2重マルコフ連鎖確率を単独で用いる場合と比較した結果を示すと図10のようになる。同図より、(i)単独タイプ2、および混合タイプ3、4、5の場合には、元の2重マルコフ連鎖確率(単独タイプ1)と同程度の文節正解率を示すが、向上効果は見られないこと、(ii)混合タイプ1、2および単独タイプ3の場合には、逆に低下していることがわかる。これは今回の実験で用いた2重マルコフ連鎖確率の値がほぼ収束の状態^{*}にあると思われ、②のような救済効果を期待することは一般に難しいと考えられる。

次に(2)の足切り効果について議論する。今回用いた2重マルコフ連鎖確率(非ゼロ確率)の中で、出現頻度数が5以下のものが全体の1割(1重マルコフの場合は、約3割)を占め、9割が50以下(1重マルコフの場合は、1,000)と大半が低頻度数である。そこで出現頻度の小さい確率を0として扱い(足切りと呼ぶ)、その部分を上記の(1)で示した組合せ方法(特に、混合タイプ3を)を使用して効果を実験により調べた。足切りの値を、10まで変化させた時の結果を、図11に示す。同図より、足切り値が1の時(足切りを行わない場合)が最良であり、足切りの効果は確認できなかった。またこのような効果は、音節のマルコフ連鎖確率を収集する時の標本データ数が少ない時にその効果が顕著に現れることが予想されるが、このときでも足切りの効果は見あたらなかった。

以上を総合すると、今回の実験で用いた2重マルコフ連鎖確率はほぼ収束の状態にあると考えられるため、これを直接にそのまま(確率が0のものは、0のまま)で用いることにより、本来の2重マルコフ連鎖確率の持つ能力が十分に発揮でき、絞り込み効果をあげることができると考えられる。

5. おわりに

本論文では、日本語処理を応用した音声認識率の向上策を検討する第1ステップとして、日本語表現の持つ音節の連鎖特性を用いた音声認識装置の音節候補絞り込みの効果を明らかにした。

実験では、新聞記事70日分の音節連鎖特性を求め、

^{*} 本来日本語として存在する音節連鎖の系列はほぼ出現しており、確率値がゼロのものは本来存在しないものとみなされることを表す。

これを用いて音節マトリックスで表された1,000文節の文節音節候補を対象に、音節候補の絞り込み効果を求めた。その結果、マトリックスの各音節候補内に正しい音節が含まれる場合は、文節音節列第10位までに、正解音節列が94.1~97.2% (第1位は60.4~66.3%)含まれ、さらに音声認識装置から出力される各音節候補の認識重み確率をも用いた場合は、10位までの文節正解率は、95.8~98.7% (第1位は72.4~78.8%)に向上することがわかった。これらの正解率は、1重マルコフ連鎖を用いた場合に比べて、10位までで18.3~21.7% (第1位で28.4~32.5%)、また認識重み確率も用いたときの10位までで5.1~9.3% (第1位で22.1~29.6%)高い値であり、2重マルコフ連鎖確率の特性が有効であることを示すものである。また本結果から、10位付近の正解率はほぼ飽和の域にあると考えられることから、3重マルコフ連鎖確率以上を用いた時の傾向としては、第1位付近の正解率をさらに大きく向上させるような効果があると推測される。

また2重マルコフ連鎖確率における基本的な問題であるマルコフ連鎖確率の組合せによる効果および足切りの効果の有無についても検討し、いくつかの代表的なタイプについて実験した。今回の範囲ではこれらの効果は認められず、ほぼ収束状態にあると考えられる2重マルコフ連鎖確率をそのまま用いる方が、音節絞り込み能力(エントロピーが小さい性質から、次に来る音節の特定効果が大きいこと)を十分に発揮できると考えられる。

今後の問題としては、音節の脱落や挿入などを含むセグメンテーション誤りにおける曖昧さの解消や、選択された音節列から漢字かな混じり文を生成する際の曖昧さを解決すること等が挙げられる。

謝辞 本研究を行うに当たり、日頃御指導いただいた基礎研究所情報科学研究部、橋本昭洋部長ならびに情報通信処理研究所自然言語処理研究部寺島信義部長ならびに、知識処理研究部村上国男部長に感謝いたします。

参考文献

- 1) 迫江, 千葉: 動的計画法を利用した音声の時間正規化に基づく連続単語認識, 音響誌, Vol. 27, No. 9, pp. 483-490 (1971).
- 2) 中津, 長島: 連続単語音声認識方式, 通研実報, Vol. 30, No. 10, pp. 2453-2465 (1981).
- 3) 岡, 鈴木: 連続 DP による音節識別に基づく連続音声認識, 音響学会音声研資, S80-20, No. 6 (1980).

- 4) 中津, 好田: 会話音声の機械認識における音響処理, 信学論, Vol. J61-D, No. 4, pp. 261-268 (1978).
- 5) 中川, 梅崎: 単音節単位に基づく音声認識法と識別された音節系列からの単語認識法, 信学論, Vol. J86-D, No. 6, pp. 1297-1303 (1985).
- 6) Raviv, J.: Decision Making in Markov Chains Applied to the Problem of Pattern Recognition, *Trans. IEEE*, Vol. IT-3, No. 4, pp. 536-551 (1976).
- 7) Jelinek, F.: Continuous Speech Recognition by Statistical Methods, *Proc. IEEE*, Vol. 64, No. 4, pp. 532-556 (1976).
- 8) 栗田, 相沢: 日本語に適した単語の誤入力訂正法とその大語彙単語音声認識への応用, 情報処理学会論文誌, Vol. 25, No. 5, pp. 831-841 (1984).
- 9) 池原, 白井: 2次3次混合型マルコフモデルによる日本語誤字訂正候補の抽出, 第31回情報処理学会全国大会論文集, 7H-5, pp. 1405-1406 (1985).
- 10) 大山, 宮井, 首藤, 吉田, 泊江: 単音節音声からのカナ文字列形成, 第26回情報処理学会全国大会論文集, 1H-6 (1983).
- 11) 坪井, 竹林, 新田, 武田: 音声による文章入力における音節接続情報の利用, 音響学会講論集, No. 3, 1-3-6 (1984).
- 12) Shannon, C. E.: Prediction and Entropy of Printed English, *Bell Syst. Tech. J.*, Vol. 30, pp. 50-64 (1955).
- 13) Kupfmüller, K.: Die Entropie der Deutschen Sprache, *FTZ*, Vol. 7, No. 6, pp. 265-272 (1954).
- 14) 齊藤: 漢字仮名混り文のエントロピー, 計量言語学, pp. 39-45 (1968).
- 15) 中野: 現代日本語における音素連続の実態 (II) 一品詞の分析, 電子計算機による国語研究V, 国立国語研究所報告, Vol. 49, pp. 81-97 (1973).
- 16) 浜田, 中津: 文字単位の音声による日本語入力法, 日本音響学会音声研究会, S84-24, No. 6 (1984).
- 17) Forney, G.D.: The Viterbi Algorithm, *Proc. IEEE*, Vol. 61, No. 3, pp. 268-278 (1973).
- 18) 宮崎, 大山: 日本語音声出力のための言語処理方式, 情報処理学会論文誌, Vol. 27, No. 11, pp. 1053-1061 (1986).
- 19) 荒木, 村上, 池原: 音韻連鎖確率による音声日本文の曖昧さの解消効果について, 第35回情報処理学会全国大会論文集, 1T-7, pp. 1327-1328 (1987).
- 20) 荒木, 村上, 池原: 2重音韻マルコフモデルによる日本語の文節音韻認識候補の曖昧さの解消効果, 信学技報, NLC 87-24 (1988).

付録 1

音節表記の日本語において, 各音節を出現確率 (0重マルコフの連鎖確率) の大きい順に並べると, 音節の出現率はほぼ指数関数的になっていると考えられる ($e^{-\lambda n}$ の形で, λ は定数, n は順位) から, 累積出現率の曲線を

$$\phi(n) = 1 - e^{-\lambda n} \quad (\text{A.1})$$

で近似させることができる。このとき各音節の確率は,

$$p(n) = \frac{d\phi(n)}{dn} = \lambda e^{-\lambda n} \quad (\text{A.2})$$

となるから, エントロピーは次式で表すことができる。すなわち自然対数を用いたときのエントロピーを H_e とすると,

$$\begin{aligned} H_e &= - \sum_{i=0}^{\infty} p(n) \ln p(n) \\ &= - \int_0^{\infty} p(n) \ln p(n) dn \\ &= - \int_0^{\infty} \lambda e^{-\lambda n} \cdot \ln(\lambda e^{-\lambda n}) dn \\ &= - \ln \lambda + 1 \end{aligned} \quad (\text{A.3})$$

通常の2を低としたエントロピーを H とすると, $H = H_e / \ln 2$ の関係および (A.3) 式より,

$$\lambda = \exp(1 - H \cdot \ln 2) \quad (\text{A.4})$$

となる。また (A.4) 式を (A.1) 式に代入すると,

$$\phi(n) = 1 - e^{-\exp(1 - H \cdot \ln 2) n} \quad (\text{A.5})$$

となる。

0重のマルコフ連鎖確率を用いたとき第1位に来る文節候補は, 各々の音節が最も高い確率を持つ時である。今平均音節候補数を4とし, 3.1節で述べたコンフュージョン表の制約 (条件①より, 各音節候補は母音ごとに平均的に26個の中から選ばれる) を考慮して, 音節候補を上記 (A.2) 式で表される音節出現率から求める。各候補の出現順位の取り方を, 1個は常に1位で残りの3個がランダムに選ばれる場合, $n=1, 7, 14, 20$ を考えると, 正規化係数 c (これは次の条件式:

$$c \cdot \sum_{i=1,7,14,20} p(i) = 1$$

より決定される) を用いて第1位の文節正解率 (これを $p\text{-}0\text{-bun}(1)$ と書くと) は, 次のように求まる。すなわち, 第1位の文節候補は, 文節内のすべての音節候補が最も高い確率を持つ時であるから,

$$p\text{-}0\text{-bun}(1) = \{c \cdot p(1)\}^L \quad (\text{A.6})$$

と表すことができる (L は文節当りの音節列長)。

次に1重マルコフ連鎖確率または2重マルコフ連鎖確率を用いた場合の第1位の文節正解率を考えると、それはそれぞれ常に前の1音節または2音節が正しい(最も高い確率)時であり、ここで前音節に依存せずに音節の出現分布関数がすべて同じであると仮定すると、1重および2重マルコフの場合の正解率 p-1-bun(1) または p-2-bun(1) も、(A.6)式で表すことができる。

付録 2

0重, 1重, 2重マルコフ連鎖確率を各々 p_0, p_1, p_2 と表す。また現在の音節 x_3 に対する前方音節 x_1, x_2 を固定した時, x_3 の値を変化させた時の最小な2重, 1重マルコフ確率を, 各々 $p_{2\min}(x_3|x_1, x_2), p_{1\min}(x_3|x_2)$ と表す。

(1) 2重マルコフ連鎖確率単独タイプ

これは p_2 が0の音節の確率値に対して, 0でない他の音節の2重マルコフ連鎖確率 p_2 の値を割り当てるタイプである。

- ① 単独タイプ 1: (従来のもの)
- ② 単独タイプ 2: $p_2(x_3|x_1, x_2)=0$ となる p_2 に対して一律に, 0でないすべての p_2 の中の最小値を割り当て, その他は①に同じ。
- ③ 単独タイプ 3: ②において, 特に同じ前方音節 x_1, x_2 に対し, x_3 と異なるある音節 x'_3 で $p_2(x'_3|x_1, x_2)=0$ の場合にのみ $p_{2\min}$ を割り当て, その他は②に同じ。

(2) 2重マルコフ連鎖確率混合タイプ

これは p_2 が0の音節の確率値に対して, 0でない他の音節の p_2 , または同じ音節の p_1, p_0 をそのまま, または正規化して割り当てるタイプである。

- ① 混合タイプ 1: 現音節 x_3 に対し, 常に $p_2(x_3|x_1, x_2), p_1(x_3|x_2), p_0(x_3)$ の中で, 最大の値を割り当てる。
- ② 混合タイプ 2: 0でない p_2 に対してはそのまま p_2 を用い, 0の $p_2(x_3|x_1, x_2)$ に対してのみ $p_1(x_3|x_2)$ を割り当てる。
- ③ 混合タイプ 3: ②の場合で, 特に0となる p_2 の中で, 同じ前方音節 x_1, x_2 に対して任意の x_3

* 同じ前方音節 x_2 において, x_3 の値を変化させた時の $p_1(x_3|x_2)$ の中で, 最大のものを $p_{1\max}(x_3|x_2)$ とする時, 正規化した値 p は, 次のように与えられる。

$$p = p_{2\min}(x_3|x_1, x_2) \cdot p_1(x_3|x_2) / p_{1\max}(x_3|x_2)$$

で $p_2(x_3|x_1, x_2)=0$ の時は0の値を割り当て, ある x'_3 で $p_2(x'_3|x_1, x_2)=0$ の時は正規化*した p_2 を割り当て, その他は②に同じ。

- ④ 混合タイプ 4: ③において, 特に0となる p_2 の中で同じ前方音節 x_1, x_2 に対して, 任意 x_3 で $p_2(x_3|x_1, x_2)=0$ の時は常に, $p_1(x_3|x_2)$ を割り当て, その他は③に同じ。

(昭和63年10月4日受付)

(平成元年1月17日採録)



荒木 哲郎 (正会員)

昭和46年福井大学工学部電気卒業。昭和51年東北大学大学院博士課程修了。同年日本電信電話公社入社。以来通信プロトコルのコンフォーマンス試験, 音声入力のための自然言語処理に関する研究に従事。現在 NTT 情報通信処理研究所自然言語処理研究部主幹研究員, 工学博士, IEEE 会員。



村上 仁一

昭和59年筑波大学第3学群基礎工学類卒業。昭和61年同大学修士課程理工学研究科理工学専攻修了。同年日本電信電話(株)入社。学生時代, 最大エントロピー法の研究に従事, 入社後音声入力における自然言語処理に関する研究に従事。現在 NTT 情報通信処理研究所自然言語処理研究部研究主任, 電子情報通信学会, 日本音響学会各会員。



池原 悟 (正会員)

昭和19年生。昭和42年大阪大学基礎工学部電気工学科卒業。昭和44年同大学院修士課程修了。同年日本電信電話公社入社。電気通信研究所, 横須賀電気通信研究所を経て, 現在, NTT 情報通信処理研究所自然言語処理研究部勤務。数式処理の研究, 待ち行列理論の研究を経て, 現在, 自然言語処理の研究に従事。大阪大学工学博士, 昭和57年情報処理学会論文賞受賞。電子情報通信学会会員。